DOCUMENT RESUME

ED 133 333 TM 005 566

AUTHOR	Spaner, Steven D.
TITLE	What Inferences Are Allowable With a Significant F in Regression Analysis?
PUB DATE	[Apr 76]
NOTE	15p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Analysis of Variance; *Hypothesis Testing; *Multiple Regression Analysis; *Tests of Significance
IDENTIFIERS	Statistical Inference

ABSTRACT

The inferences allowable with a significant F in regression analysis are discussed. Included in this discussion are the effects of specificity of the research hypothesis, incorporation of covariates, directional hypotheses, and the manipulation of variables on the interpretation of significance for such purposes as causal and directional inferences. The position is taken that the research hypothesis dictates the variables to be tested and hence included in the regression models. If the variables have been manipulated, then causal inferences can potentially be made. If covariates have been included in the analysis, then they should be included in the inference. If the research hypothesis specified a directional expectation, then a directional conclusion is warranted. (RC)

/.2/

ED133333

What inferences are allowable with a

significant F in regression analysis?

Steven D. Spaner

University of Missouri - St. Louis

A paper presented at the symposium "Issues in the use, interpretation, and teaching of multiple linear regression" sponsored by the Special Interest Group on Multiple Linear Regression and the American Education Research Association Convention,

San Francisco, 1976

7.21

TM005 566

What inferences are allowable with a significant F in regression analysis?

Steven D. Spaner

University of Missouri - St. Louis

Multiple linear regression (MLR) procedures (Wainer, 1976) and the linear model in general (Brown, 1975) have come under attack in the most recent months. And, sad to say many of the points raised are cogent criticisms indeed. However, none of the remarks are foreign to the members of SIG-MLR or the readers of Multiple Linear Regression Viewpoints. But, perhaps these admonitions from our brethren would be less audible if the topics to be discussed here today were more universally heard. What appears to be needed is a better understanding of MLR: its foundations, its applications, and its ramifications. My presentation addresses some of these misunderstandings.

We calculate regression coefficients, we determine R^2s, we compute F statistics, but do we know what it's all about? If the audience is composed of Ward-Kelly-McNeil protégés then the word "assumptions" elicits a discounting laugh; but there are those who read a list of assumptions in a statistics text and live by them or die by them. I believe an examination of the foundations of MLR would do both the jester and the jouster some good.

Drawing from a variety of sources (Ward & Jennings, 1973; Lindquist, 1953; McNeil, Kelly, & McNeil, 1975; Snedecor, 1956; Young & Veldman, 1972; and Glass & Stanley, 1970) the following list of assumptions have been identified for the F statistic:

1. The F statistic must have been generated from randomly selected and independent entities or criterion measures.
2. The variance of the criterion measures within each population subgroup must be equal (homogeneity of variance).
3. The distribution of the criterion measures in each population must be normal.

Specific authors have added to or restated in other words the above list. The only meaningful change would occur in the case where a <u>covariance analysis</u> had been performed and the F statistic generated from these data. In such an instance the above assumptions apply to the adjusted criterion measures and a fourth assumption is added:

4. The regressions of the criterion measures onto the covariate(s) are equal for each population subgroup (homogeneity of regression).

Before I summarily dismiss these F ratio assumptions with an impressive list of citations, it should be pointed out that another list of assumptions are of concern. While the F distribution is the theoretical sampling distribution, the overall calculation technique is regression, which is synonymous with correlation (in point of fact, Galton postulated his "law of universal regression" before Pearson developed the index of correlation/ship/). Hence we must also acknowledge the set of assumptions underlying regression. These assumptions are essentially the characteristics of the bivariate normal distribution:

1. X scores, disregarding Y scores, are normally distributed.
2. Y scores, disregarding X scores, are normally distributed.
3. The Y scores for each X score are normally distributed with a common variance ($\sigma_{y.x}^2$).
4. The X scores for each Y score are normally distributed with a common variance ($\sigma_{x.y}^2$).

5. The means of the Y score distributions for each X score fall on a straight line.

Since we are here to discuss multiple regression the above assumptions must be expanded to address the multivariate normal distribution; but I will leave that task to the reader, the features are essentially the same. However, the expansion brings up my first point with regard to allowable inferences.

Snedecor (1956) has classified multiple regression into two basic models: Model I - the values of X are considered fixed, that is, chosen by the investigator, only the Y values or the criterion is a random, normally distributed variable, and Model II - the values of X are not selected, individuals are randomly selected leaving the values of every variable measured on the individuals available to chance; that is, a random sample is drawn from a multivariate normal population. The first model is exemplified by the *ex post facto* research design and designs in which treatments are not randomly assigned. The second model is more in the tradition of the experimental design. Investigations utilizing model I often take the form of covariance analyses which call for the 4th assumption to the F statistic and call for a limitation on the generalizations; i.e., the population of adjusted Y scores. However, statistical control should not be viewed as limiting since it allows one to study the actual situation instead of one that has been artificially produced by experimental control.

Lindquist (1953) has observed that in educational research the application of model two is often amended to random assignment of treatments instead of random assignment of subjects to treatments. The reasons

are obvious and the amendment ingeniously adaptive; however, it must be understood by the researcher that his population has changed. The population is no longer individuals who are potentially available for random selection but instead intact groups who are available for random assignment of treatments.

Lest you begin thinking that assumptions are all this paper is going to address let me put your thoughts at rest. A number of investigations have dealt with the F distribution assumptions and their violation (Norton, 1952; Bonneau, 1960, 1963; Young & Veldman, 1963; Pearson, 1931; Box & Anderson, 1955). The summarized conclusion from these investigations is that there is no appreciable effect on the accuracy of the F test from nonnormality and if sample sizes are equal, heterogeneity of variance has a negligible effect. The only apparent serious violation that can be committed is failure to randomly select independent entities or measures (Glass & Stanley, 1970). However, surprisingly, there are no impirical investigations of this tenet. Furthermore, there is a dearth of inquiry into the effects of violation of the homogeneity of regression assumption and the list of postulates for regression and correlation. From the few statements offered on these topics (Snedecor, 1956; Vasu & Elmore, 1975) it appears that once again normality is a mute issue but that dependence of observations ($r \geq .95$) can cause disruption of accurate calculations. Snedecor (1956) has recommended the elimination of one of the pair of Xs with correlation greater than .95 (based on a redundancy interpretation). However, McNeil and Spaner (1971) have made a case for a more judicious examination of such a recommendation, especially as it would apply to nonlinear problems.

Having fulfilled my obligation to the conservatives; having acknowledged the assumptions we are operating under in MLR; I will now turn to the topic sentence: "What inferences are allowable with a significant F in regression analysis?" The first and foremost limitation on our inferences is the research tool; i.e., the type of regression technique, stepwise or hypothesis testing regression. My colleague (McNeil, 1976) has addressed himself to this matter so I will make only a brief remark. The calculation of an F statistic carries with it the implication that a comparison is desired, a decision is to be made, and that an inference will follow. All three of these activities suggest that a comparison hypothesis has been adopted (a null hypothesis) and an alternative hypothesis will be accepted should the comparison hypothesis be found untenable. Hypothesis testing regression fits this research format precisely: a null hypothesis is formulated - the restricted model as Bottenburg and Ward (1963) tagged it, and an alternative hypothesis is proposed - the full model in Texas terminology (this form of labeling has been principally associated with University of Texas faculty and graduates). A significant F statistic, in the "grand tradition", calls for rejection of the null hypothesis (restricted model) and acceptance of the alternative hypothesis (full model).

Let us look now at stepwise regression. Stepwise comes in two forms: ascending and descending. Ascending stepwise regression adds variables to a null set until a new set is created which has maximum predictive efficiency (according to some criterion). Descending stepwise regression operates in just the reverse; from a defined set of variables, variables are removed which least add to efficient prediction (until some "stop" criterion is obtained). Examination of these two techniques in

relation to an F test leaves the researcher with a "loose end". In the case of ascending stepwise, the null hypothesis is known, it's a model of the criterion grand mean. But, the alternative hypothesis is unknown; hence, no decision can be contemplated and no inferences entertained that are not sample generated. Likewise, in descending stepwise the alternative hypothesis is known, but we must wait for the "marvelous toy" to tell us what our comparison (null) hypothesis is. Therefore, if inference and generalizations are to be allowed upon significant F tests we must conduct hypothesis testing regression analyses.

A second limitation on our allowable inferences relates to the research hypothesis. Many of the more interesting and pertinent questions in education call for the statistical control of variables that are practically or explicitly beyond experimental control. Indeed, one of the attractive features of MLR is the ease with which covariance analysis can be conceptually as well as operationally understood (Williams, 1976). And while we have a number of admonitions against causal interpretations in ex post facto (basically correlational) studies (Campbell & Stanley, 1969; McNeil, Kelly, & McNeil, 1975; Newman, et. al., 1976); we would do well to remember, ". . . that in many cases statistical control is more to be desired; the actual situation is studied instead of one artificially produced, the observations are extended over a greater range, thus broadening the foundation for inference, and in the end one has knowledge of the variation of two [or more] quantities instead of one, together with the relationship between them." (Snedecor, 1956, p. 146) However, covariance analysis it will be remembered imposes a 4th assumption on the F distribution - homogeneity of regression. Ironically, McNeil and company (1975, p. 131) have shown us how easy it is to make a test of this assumption with MLR

(i.e., interact the membership variables with the covariates and test the equality of the coefficients); yet, users and teachers (Williams, 1976) seem to be unfamiliar with this easy but necessary test. Hence, inferences based on significant F tests of covariance analyses will be rendered inaccurate if not invalid without the homogeneity of regression test.

Indeed the value of and need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding. True, a significant interaction hampers the interpretation of main effects, but the positive view is that a significant F test of interaction tells us how to appropriately limit our generalizations (Glass & Stanley, 1970).

Another issue related to hypotheses and having a bearing on our inferences with a significant F statistic is that of directionality. Without going into the rudiments of sampling theory, I will summarize the directionality vs. nondirectionality decision as one which doubles your chances of rejecting the null hypothesis, if you hypothesize in the right direction. It's that last phrase that seems to be overlooked by many researchers. There seems to be a preponderance of cautious research hypotheses (nondirectional) and bold research conclusions (directional). Let me give some guidance as to this choice of hypotheses as it relates to MLR. It must be remembered that regression, especially least squares regression, is basically a curve fitting technique (Lewis, 1960; Snedecor, 1956). This being the case there are only three aspects of a curve that are manipulatable: 1) the point where the curve intersects with some reference axis (intercept point), 2) the rate of rise of the curve (the slope, which is an indicator of relationship), and 3) the number of inflection points

in the curve (this is governed by the exponentiation factor). Of these three factors only one is available for nondirectional hypotheses: the comparison of intercept points. The weightings or regression coefficients that identify level of intercept are indicative of group means. And direction of differences between group means can be unknown such that a nondirectional hypothesis is conceivable. Of course, each researcher is characterized by his or her own risk-taking-behavior, but there are strong arguments for stating all research hypotheses as directional hypotheses (McNeil, Kelly, McNeil, 1975).

The other two aspects of a curve call for directional hypotheses by the very nature of their source. The source of rate or slope changes is the addition of information into a model. The worse that can happen is no change; hopefully the added information will enhance the relationship of the predicted scores with the actual scores (i.e. increase $r_{y\hat{y}}$). And, since our yardstick of measurement is error sum of squares we cannot generate more prediction error by including information which is related to the criterion. We can, however, possibly reduce the error by the inclusion of new information (Ward & Jennings, 1973; Snedecor, 1956; McNeil, Kelly, & McNeil, 1975).

The same mathematical truths hold for hypotheses about inflection points. That is, in hypothesizing models that are to fit nonlinear data, the only possible hypothesis is a directional hypothesis. Inclusion of exponentiated variables allows the best fit line to bend and turn with the data thus reducing the sum of squared deviations from the line. However if the exponentiated variables do not create a better fit, they do not create <u>more</u> error, they simply take up degrees of freedom without error reduction and receive zero or low weightings.

Hence, the point to be made is that only directional hypotheses are being tested with a significant F in two of the types of regression

hypotheses, whether the researcher has stated them as such or not. Only in the statement of hypotheses about intercepts (group mean differences) is there the potential for a nondirectional F test and there is great suspect of a researcher who would not have some expectation of the direction of mean differences.

Returning momentarily to a point made about nonlinear fits of data; there are recurring calls for the development of nonlinear hypotheses in educational research, the most recent coming from Brown (1975). Being myself one of the more ancient heralds of this idea (McNeil & Spaner, 1971) and knowing that I was not the first to send smoke, I find it a puzzlement as to why there are not more nonlinear hypotheses in educational research. I come up with two possible reasons: one, the "Pandora's box" fear, and two the fear of violation of assumptions. With regard to the first reason, Pandora's box, it is true that there is an infinite set of nonlinear terms. However, bivariate plots of sample data can narrow down the field of fruitful hypotheses quite well. Additionally, there are computer aids to "snoop around" in sample data and arrive at some tenable hypotheses (Automatic Interaction Detection-Version 4 by Koplyay, Gott, & Elton, 1973).

The second fear, violation of assumptions, must refer to the linearity of regression assumption since the first portion of this paper has indicated that normality and homoscedasticity are not critical. So let's examine the meaning of this assumption: it states that the means of all sampled populations lie on a straight line. But what is the effect of this assumption? It allows us to infer for populations not sampled. We assume the population mean for unsampled populations falls on the sampled regression line. And, this is a valuable principle, indeed, for without

it we have no prediction. However, does this regression line have to be straight? If it does, then there are a number of logarithmic transformations of curve line functions that will bring the curved best fit line back into line (Lewis, 1960). But all this manipulation is unnecessary; it is simply a rescaling process. It's a recognition that the measuring devices we use are not necessarily monotonic. It's a recognition, as Pohlman & Newman (1973) have suggested, that the assumption of rectilinearity has not been met. But more than that, it's a recognition that the assumption can be met if a curved best fitting line is used. McNeil & Kelly (1970) and Ward & Jennings (1973) have addressed themselves to this issue and have suggested that the investigator attempt to express functional relationships in data first and not worry about whether conditions or assumptions are met. In other words, if it works, use it.

As a concluding remark, it should be pointed out that "what inferences are allowable with a significant F in regression analysis "are zero if the R^2 of the regression model is not practically significant." Ward & Jennings (1973) have indicated that very large samples can produce statistically significant F statistics with very little practical consequences. Scientific rigor and the scientific process produce significant results, statistics simply apply a probability level to your potential error in judgment. Focus should be placed on 1) random and independent sampling, 2) clear and precise statements of research hypotheses, 3) construction of appropriate models to test the hypotheses, 4) examination of R^2 values, and 5) allowable inferences from significant findings.

References

Boneau, C.A. The effects of violations of assumptions underlying the t-test. Psychological Bulletin, 1960, 57, 49-64.

Boneau, C.A. A comparison of the power of the U and t-test. Psychological Review, 1963, 69, 246-256.

Bottenberg, R.A. & Ward, J.H. Applied multiple linear regression (AD 413128). Lackland Air Force Base, Texas: Aerospace Medical Division, 1963.

Brown, D.J. Mirror, mirror . . . . down with the linear model. American Educational Research Journal, 1975, 12, 491-505.

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago, Illinois: Rand McNally, 1969.

Glass, G.V. & Stanley, J.C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Koplyay, J.B., Gott, C.D. & Elton, J.H. Automatic interaction detection-version 4 (AID-4) (AFARL-TR-73-17). Lackland Air Force Base, Texas: Personnel Research Division, 1973.

Lewis, D. Quantitative methods in psychology. New York City: McGraw Hill, 1960.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston, Massachusetts: Houghton Mifflin Co., 1953.

McNeil, K.A. Position statement on the roles and relationships between stepwise regression and hypothesis testing regression. Multiple Linear Regression Viewpoints, 1976, 6, 46-49.

McNeil, K.A. & Kelly, F.J. Express functional relationships among data rather than assume "intervalness." The Journal of Experimental Education, 1970, 39, 43-48.

McNeil, K.A., Kelly, F.J. & McNeil, J.T. Testing research hypotheses using multiple linear regression. Carbondale, Illinois: Southern Illinois University Press, 1973.

McNeil, K.A. & Spaner, S.D. Brief report: highly correlated predictor variables in multiple regression models. Multivariate Behavioral Research, 1971, 6, 117-125.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R. & Ervin, L. Type VI error: inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 1976, 6, 1-19.

Norton, D.W. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. An unpublished dissertation, 1952. Summarized in Lindquist, E.F. Design and analysis of experiments in psychology and education (pp. 78-90). Boston, Massachusetts: Houghton Mifflin Co., 1953.

Pohlman, J.T. & Newman, I. Regression effects when the assumption of rectilinearity is not tenable. Multiple Linear Regression Viewpoints, 1973, 4, 51-59.

Roscoe, J.T. Fundamental research statistics for the behavioral sciences (2nd ed). New York City: Holt, Rinehart & Winston, 1975.

Snedecor, G.W. Statistical methods (5th ed). Ames, Iowa: Iowa State University Press, 1956.

Vasu, E.J. & Elmore, P.B. The effect of multicollinearity and the violation of the assumption of normality on the testing of hypotheses in regression analysis. Multiple Linear Regression Viewpoints, 1975, 6, 21-50.

Wainer, H. Estimating coefficients in linear models: it don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Ward, J.H. & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

Williams, J.D. Should a first course in ANOVA be through MLR? Multiple Linear Regression Viewpoints, 1976, 6, 42-45.

Young, R.K. & Veldman, D.J. Heterogeneity and skewness in analysis of variance. Perceptual and Motor Skills, 1963, 16, 588.

Young, R.K. & Veldman, D.J. Introductory statistics for the behavioral sciences (2nd ed). New York City: Holt, Rinehart, and Winston, 1972.